Evo Edu Outreach (2012) 5:501–507
DOI 10.1007/s12052-012-0436-8

CURRICULUM AND EDUCATION ARTICLE

# Where Do I Come From? Using Student's Mitochondrial DNA to Teach About Phylogeny, Molecular Clocks, and Population Genetics

**Luana S. Maroja · Jason A. Wilder**

Published online: 5 September 2012


**Abstract** Phylogenetic reconstruction, divergence times, and population genetics are critical concepts for a complete understanding of evolution. Unfortunately, students generally lack "tree-thinking" skills and are often unmotivated to explore these concepts using typical classroom exercises that feature taxa unknown to students or simulated datasets. To generate greater student interest, we have developed an affordable practical lab ($16 dollars per student) where students extract and sequence their own mtDNA and use it for exercises involving phylogenetic reconstruction (placement of own DNA into the world tree), divergence (speciation) time (comparing current student population with chimps, gorillas, and Neanderthal), and population genetics (demographic change calculation based on student's sample). In contrast to traditional labs, we found that students were highly motivated and enthusiastic throughout the four-week activity. Students had a 100% rate of success in obtaining DNA sequences and their evaluations report high satisfaction with the learning outcome. Here we provide all details and datasets needed to run the lab and discuss a series of assessments and possible exercises.

**Keywords** Tree-thinking · Gene genealogies · Population genetics · Phylogenetic analysis · Phylogeography · Human evolution

**Electronic supplementary material** The online version of this article (doi:10.1007/s12052-012-0436-8) contains supplementary material, which is available to authorized users.

L. S. Maroja (✉)
Department of Biology, Williams College,
Williamstown, MA 01267, USA
e-mail: lsm1@williams.edu

J. A. Wilder
Department of Biological Sciences, Northern Arizona University,
Flagstaff, AZ 86011, USA
e-mail: jason.wilder@nau.edu

## Introduction

"Tree-thinking" is an essential concept in modern biology and a necessary tool in uncovering evolutionary relationships (Sandvik 2008; Baum et al. 2005; Baum and Offner 2008). Yet tree-thinking is still virtually absent among students ranging from non-majors to graduate students (Sandvik 2008; Meisel 2010; Gregory 2008). Fostering skills in tree-thinking is not only an essential component in biological education, but it also helps integrate evolutionary concepts throughout the curriculum (Baum and Offner 2008; Offner 2001). Although every student taking an introductory biology or non-major course should develop some basic tree-thinking skills (Baum et al. 2005), biology majors should also develop a basic understanding of the technical details behind tree-building (i.e., how phylogenies are inferred by practicing systematists). While the non-technical tree-thinking skills (i.e., interpretation of phylogenetic trees) can be taught in more traditional lecture/assignment styles (Baum et al. 2005; Baum and Offner 2008; Gregory 2008), the practical aspects of tree building require practice, the use of technical computer programs, and access to molecular datasets. In typical exercises, these datasets are often either artificial or are downloaded from NCBI's Genbank and are comprised of taxa unknown or of little interest to students. Based on our experience and course surveys, this method often fails to motivate most students and can result in poor learning of both technical and basic tree-thinking skills.

Students are more highly motivated and learn best when they are personally interested in the final outcome of a project (Handelsman et al. 2007). One of the ways to make students interested in the final results is to involve them personally with the exercise. Recently, several excellent techniques have been developed to address tree-thinking problems (Baum et al. 2005; Gregory 2008; Smith and Cheruvelil 2009), and the laboratory-

teaching unit developed here adds an additional approach that gets students directly involved with results and outcome. In this laboratory-teaching unit, students use their own DNA to place themselves in the humans of the world mtDNA phylogenetic tree (Fig. 1) and to deduce and find the geographic origin of their maternal lineage and to estimate population genetic parameters. They then use their human sequences and that of other species of apes and Neanderthals to examine evolutionary relationships among species and to calculate substitution rates.

Here we describe a four-week laboratory-teaching unit and its learning goals, activities, and assessment. The activities include wet labs (DNA extraction, PCR, and sequencing), sequence analysis (cleanup, alignment), phylogenetic reconstruction (tree building and molecular clocks), and population genetic analysis (genetic diversity, demographics). These labs were designed for an upper-level biology major class (we use it to accompany a junior-level evolutionary biology course), but this can easily be adapted for introductory classes or even for non-major courses. The overall cost per student was $16 dollars in consumables (price estimated in 2012).

## Concepts Addressed

This laboratory-teaching unit addresses most aspects of phylogenetic tree reconstruction and some population

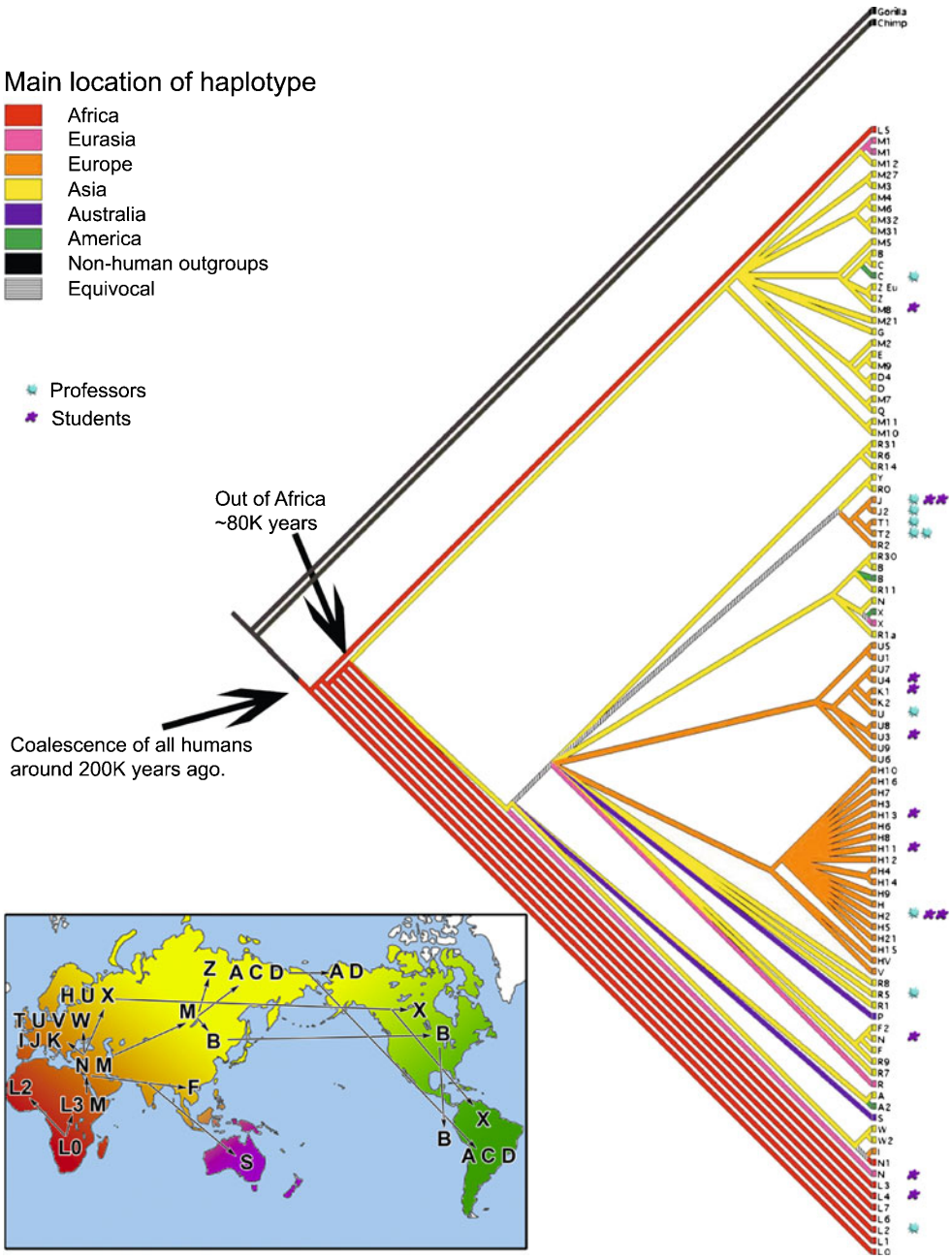


**Fig. 1** World cladogram built with complete mtDNA of known haplotypes (for color see online version). Tree was built with Mr. Bayes (Huelsenbeck and Ronquist 2001), GTR+I+G model, five million generations, and visualized in MacClade 4.0 (Maddison and Maddison WP. MacClade 4: analysis of phylogeny and character evolution. Version 4.08a. http://macclade.org 2005). Students and professors that were successfully assigned to haplotypes (70% of sequences clustered to a known haplotype with a posterior probability higher than 65%) are plotted as *dots* on tips corresponding to their mtDNA haplotypes. A simplified migration map (*inset*) shows the likely path of migration and origin of main haplotypes. This figure has great educational value, helping students understand how phylogenetic data can be used to reconstruct migration patterns (ancestral haplotypes of African origin and out-of-Africa migrations, low genetic diversity in Americas indicating recent colonization). Major events in human evolution are shown in *arrow*, but the tree is not to scale. Data in Nexus format used to build this tree can be found in supplementary materials (ESM 4) and accession number in supplementary materials Table 1

genetic and molecular evolution concepts. At the end of this exercise, students will have a solid understanding of how data are collected for phylogenetic reconstruction, encompassing the entire process from DNA extraction to data analysis. The first two weeks of lab are fully practical, including DNA extraction, PCR, gel visualization, PCR cleaning, and sequencing (see the "Materials and Methods" section). The third week is focused on data processing (cleanup and alignment) and phylogenetic analysis, and the final week focuses on molecular clocks, gene genealogies, and some population genetic analysis.

The data collected by students can be used to address many of the misconceptions involved in phylogenetic reconstruction and analysis (recently reviewed in Meisel 2010; Gregory 2008). For example, using data from extant human populations can help students understand that basal mtDNA clades are not "primitive" and, in fact, comprise an important component of modern human diversity. Moreover, the mtDNA genealogy illustrates concepts such as the MRCA (Most Recent Common Ancestor) both in terms of the ultimate basal node of the tree and for individual clades, such as that comprising the out-of-Africa component of human history. In addition, these data can help students understand how patterns of migrations can be inferred from gene genealogies (phylogeography and coalescence). As an example, the American populations are not only lacking in genetic diversity (indicating recent colonization by few people), but are also directly related to the Asian haplotypes, confirming the anthropological evidence of migration from Asia through the Bering Strait. Using the freeware program MEGA (Tamura et al. 2011) to examine the world mtDNA tree, students can understand the concept of roots and clades, different ways to represent a tree, and how rotating internal nodes does not change the tree topology or evolutionary history. With these kinds of exercises, students can get around the common misconceptions of the "great chain of Being" (Gregory 2008) in which living species (in this case, individuals) are ranked lowest or highest and the "main line and side tracks" misconception (Gregory 2008; Omland et al. 2008) where evolutionary history is interpreted as a progressive process with a superior distinctive "end point" at the end of the main line.

Finally, the student-collected data can be combined with NCBI Genbank data of other species such as Neanderthals, great apes, and old world monkeys. With these data, questions about the molecular clock and species relationships can be explored. With the freeware DNAsp (Librado and Rozas 2009), students can estimate population genetic and demographic parameters and also develop a deeper understanding of the difference between gene genealogies and population history (i.e., that gene genealogies coalescence might predate fossil age and speciation).

### Learning Goals

We had three overarching goals, encompassing the entire series of steps that would be performed by a practicing systematist or population geneticist.

- To reinforce lab techniques superficially learned in introductory biology courses, such as DNA extraction, PCR, and sequencing
- To help biology majors develop their tree-thinking skills and their basic understanding of the technical details behind tree construction from data collection to analysis
- To help students develop a deeper understanding of gene genealogies, molecular clocks, and basic population genetics parameters

## Materials and Methods

### First Lab: DNA Extraction, Quantification, and PCR Set Up. Estimated Time, About Two Hours

Prior to DNA extraction, procedures were implemented to ensure complete anonymity of student samples. In addition, students were given the option of not processing their own DNA and instead using DNA from an anonymous non-student donor (collected beforehand from willing faculty and staff). Each student was given a random letter/number code only known to them and was also provided with anonymous cheek cells which they could choose to process in place of their own sample. It is important to note that educational activities, even those involving human subjects, do not constitute "research" and therefore do not require IRB (institutional review board) approval. However, we encourage educators considering this activity to contact their local IRB before the course's onset to ensure adequate oversight and compliance with local procedures. Indeed, participating in a human subject training course (offered by most institutional IRBs) prior to implementation of this activity will help to raise instructor awareness of the types of issues that may arise when collecting data in a classroom setting (this information can form the basis of interesting discussions with students as well).

Students extracted DNA from cheek cells (harvested with a sterile OmniSwab) using a QIAamp DNA Mini Kit (Qiagen), with a final elution in 150 μl of AE buffer. After extraction, DNA was quantified with a Nanodrop 2000 (ThermoScientific) [Note: for large classes, the quantification step could be skipped, all extractions contained DNA (from two ηg/μl to 20 ηg/μl) and

all PCRs were successful]. If a Nanodrop or similar equipment is not available, an alternative method is agarose gel DNA quantification with or without a DNA mass ladder (such as "ThermoScientific Fermentas MassRuler™"); using at least ten μl of DNA will allow visualization of even the lowest concentration samples (however even samples that cannot be visualized will often yield successful PCR products).

The entire hypervariable region was amplified in a single PCR step (Fig. 2, primer information in Table 1). To minimize the chance of error, a master mix was prepared before class so students only had two pipetting steps (mixing their DNA to the PCR master mix). PCR conditions and cycling protocol can be found in Table 2. In our course, each student prepared three PCRs, one with their own DNA, one with DNA of an anonymous donor, and a control PCR with water in place of DNA.

Notice that due to the position of the origin in the middle of the hyper variable region (D-loop), HVR2 is actually located at the start of the mtDNA molecule (position 1–574). Thus the collected sequence data will have the order of HVR1 and HVR2 inverted (Fig. 2). The acquired sequences will start at HVR1 and finish in HVR2. Thus if other sequences from NCBI Genbank are added to the dataset, the order of HVR1 and HVR2 must be reversed for proper alignment (the Nexus files provided in supplementary material ESM 5 are already in proper order, HVR1 followed by HVR2).

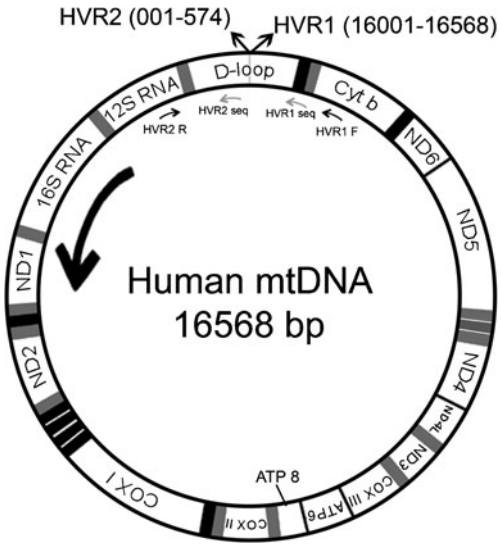


**Fig. 2** The human mtDNA genome, totaling 16,568 nt pairs with the numbering starting inside the D-loop (hypervariable region) and proceeding counterclockwise (*large arrow*). In addition to 13 genes named in the figure, the mtDNA encodes two rRNAs (12S and 16S) and 22 tRNAs (in *black* and *gray stripes* representing tRNAs transcribed from the L-chain and H-chain, respectively). PCR primers are indicated as *small black arrows* inside the mtDNA and sequencing primers as *gray arrows*

**Table 1** PCR primers and sequencing primers (SEQ) used to amplify the entire human hyper variable region, see Fig. 2

| Primer name, direction (F or R), and position in relation to human complete mtDNA | Sequence (5′-3′) |
|---|---|
| HVR1 F (15614-15637) | AGG CGT CCT TGC CCT ATT ACT ATC |
| HVR2 R (767-744) | CGT GCT TGA TGC TTG TCC CTT TTG |
| HVR1 SEQ F (15986-16009) | CCA TTA GCA CCC AAA GCT AAG ATT |
| HRV2 SEQ F (104-127) | CCG GAG CAC CCT ATG TCG CAG TAT |

## Second Lab: Gel Visualization and Preparation for Sequencing. Estimated Time, Two to Three Hours

Agarose gels at a concentration of 1.5% with TBE buffer and two μl of GelRed (Phenix, a safe alternative to Ethidium bromide) in 50 ml of gel were pre-casted before lab (alternatively students could cast their own gel, adding 40–50 minutes of lab time). Each gel containing ten wells was shared by three students (three PCRs each); the first lane was loaded with one kb+ ladder (5 Prime XL ladder). Students mixed their PCRs (student, professor, and control) with loading dye on separate tubes prior to gel loading. Gels were run for 40 minutes and visualized on UV light with photo capacity. After confirmation of correct size product, students loaded five μl of a master mix consisting of two μl of Exonuclease I and three μl Shrimp Alkaline Phosphatase (EXO/SAP Thermo Scientific Fermentas) to each tube. Reactions were incubated for 50 minutes at 37 °C

**Table 2** PCR conditions and cycling protocol with Perfect*Taq* DNA polymerase™ (5 Prime)

| Components | Total volume (μl) | Final concentration |
|---|---|---|
| 10× PCR Buffer | 5 | 1× |
| dNTP mix (ten mM each) | 1 | 200 μM |
| Primers (ten μM) each, Table 1 | 1 each | 0.2 μM each |
| Water (RNase free) | 31.75 | – |
| Perfect*Taq* enzyme | 0.25 | 1.25 units |
| DNA | 10 | ~20–200 ηg |
| Total volume | 50 | – |

Initial denaturation of three min at 94 °C followed by 35 cycles of 30 s at 94 °C, 30 s at 51 °C, and 60 s at 72 °C and a final extension of ten min at 72 °C

followed by a heat deactivation of ten minutes at 90 °C (students were free to leave the lab during this time). Samples were then prepared for sequencing (direct PCR sequencing by RETROGEN, primers mixed with DNA) by adding 3.5 μl of EXO/SAP treated PCR product, 0.5 μl of primer (ten μM), and six μl of water on a well-labeled eppendorf tube (if lab time should be kept shorter, students can label tubes and professor/teaching assistants can load PCRs after EXO/SAP is completed; this reduced lab time by one hour). Each PCR was sequenced twice using the HVR1 sequencing primer and the HVR2 sequencing primer (Table 1). Samples were shipped to RETROGEN by overnight FEDEX and had a two–three-day turnover. A list of materials needed for the two wet labs is provided in supplementary material table 3.

Third Lab: Visualization and Clean Up of Sequence, Alignment of Students' Sequences with Known Haplotypes, Basic Phylogenetic Tree Reconstruction. Estimated Time, Two to Three Hours

During this lab, students downloaded their ".ab1" files, visualized, and aligned the two sequencing reactions (HVR1 and HVR2 sequencing primers) for each individual. To visualize and clean up sequences, we used the freeware MEGA5; instructions on how to create a single-edited sequence per student using only freeware can be found in supplementary information material ESM 3 and ESM 6. Students then aligned their DNA sequences to the previously created haplotype file (available in ESM 5; average sequence length was 1,327 bp) that contains representative sequences from all known major human haplotypes (see supplementary material Table 1 for accession numbers of each sequence). Alignment was done by eye (the world's haplotypes were already pre-aligned and it is faster and more precise to align by eye than using an alignment function). Distance-based trees can be built with Mega 5 (Tamura et al. 2011; Tamura et al. 2007) and using this feature, the students could get a rough idea of their assigned haplotypes—in about 70% of the cases, the assignment remained the same with more strict reconstruction methods (likelihood or Bayesian); however in some cases (~30%), sequences became non-resolved (basal to all others). Part of these unassigned haplotypes could be explained by the presence of indels (likely due to heteroplasmy) within highly repetitive regions (about 15% of individuals); as a consequence, part of the sequence was unreadable. In some of these cases, it was still possible to assign a haplotype; however, a few individuals were non-resolved due to lack of information (i.e., short sequence).

As a follow-up exercise (which can be done as an assignment), students can export their data (and world's haplotypes) as a Nexus file and run a rigorous model-based tree search in Mr. Bayes (Huelsenbeck and Ronquist 2001; five million generations, GTR+I+G; alternatively, this step can be done by the professor or teaching assistant). Model-based tree searches often resulted in a loss of resolution (especially for short sequences) in relation to the distance matrix methods. Students should understand that this is in part because distance matrix methods are unable to account for homoplasies (common in fast-changing regions such as the HVRs); distance matrix methods just measure the distance between each pair of species, leaving out all information from higher-order combinations of character states (Felsenstein 2004).

Fourth Lab: Population Genetics, Molecular Clocks, and Speciation Time. Estimated Time, Two to Three Hours

For the final lab, students used the free software DNAsp (Librado and Rozas 2009) to analyze a dataset (available in ESM 7) consisting of student sequences, professor sequences, *Homo sapiens sapiens* Neolithic sequences (of various ages), *Homo sapiens neanderthaliensis* sequences (of several ages), *Pan paniscus* (bonobo), *Pan troglodytes* (chimp), and *Pongo pygmaeus* (orangutan) sequences (complete data file was aligned before class and provided to students). Because only HVR1 was available for most of these sequences, the region used was smaller than that used for haplotype assignment and phylogenetic reconstruction (only 402 bp of 1,327 bp were used). Accession numbers of sequences used are listed in supplementary Table 2 (Suppl_inf_tables).

## Assessment

The third and fourth labs of this teaching unit provide opportunities for formative assessment (assessment during the teaching unit (Handelsman et al. 2007)) in the form of discussions and exercises during class and should be followed by summative assessment (assessment at the end of unit (Handelsman et al. 2007)) in the form of a written assignment using data collected and analyzed by students (suggestions for questions can be found on "ESM 2"). During the third lab (data analysis and phylogenetic reconstruction), while students were building basic phylogenetic trees and manipulating trees in Mega (Tamura et al. 2011; Tamura et al. 2007; see "ESM 3"), they should be turning to questions related to "tree-thinking" (Baum et al. 2005; Gregory 2008; Omland et al. 2008). Our main goal during this discussion was to make students aware of main misconceptions on phylogenetic interpretation and help them understand how patterns of migration (phylogeography) could be acquired from data (students were also provided with copies of Fig. 1). We focused on the "reading along the tips ladderized

misinterpretations" and "clade density and node counting" (Meisel 2010) misconceptions. The first involves the idea that some clades are primitive while others are advanced and that "primitive" clades gave rise to more "advanced" ones, whereas the second involves the notion that the number of nodes indicates distance from a common ancestor (i.e., strait lines are not "evolving"). Taking advantage of Mega's (Tamura et al. 2011; Tamura et al. 2007) great tree visualization capacities including subtree flipping and swapping (to rotate branches), rooting tool, clock calibration, and different ways to visualize trees (topology only, traditional, circle, etc.), we emphasized two features of the tree that help to correct these misconceptions. The first is that branches can be rotated without changing the topology of the tree (thus "reading along the tips" is not the correct way to read a tree) and the second is that clade branch lengths are similar (clock-like), indicating that all sequences are evolving at a similar rate whether or not they are deep in the tree (thus "straight lines" are equally distant from the ancestor). The second misconception can also be corrected by pointing out that "straight" lines are often African clades of which subclades have not been described (for example H clade is the one with most subgroups as it represents mostly European sequences, a very well-researched group, see Fig. 1), if more sequences were available, the L clades could be substantially expanded by many added nodes. Finally, this mtDNA gene genealogy presents a good opportunity to discuss the differences between gene genealogies and species trees. It should be clear to students that we are not dealing with different species, but with individuals within *H. sapiens*. Ideas about gene genealogies and species phylogenies can be introduced: for example, the notion that not all gene genealogies show the same pattern/topology. In mtDNA, the MRCA is expected to be more recent because the effective population size is only one fourth of the genomic one (maternal and haploid), thus coalescence processes are expected to progress faster. Furthermore gene genealogies often have the complicating factor of ancestral polymorphism (which could lead to conflicting gene trees) and introgression (as observed between humans and Neanderthals (Green et al. 2010) at the genome level). By analyzing the tree, students should notice that even in this limited dataset (only a fraction of the mtDNA), it is clear that the most basal haplotypes are African and that Africa has the most divergent sequences (e.g., European sequences coalesce quickly to a common ancestor) and the most genetic diversity (students can compute this with DNAsp in the last lab), while the Americas are the most depauperate continents with only a handful of closely related haplotypes. The data thus support the idea that Africa was the place of origin for humans, with a population that had time to accumulate genetic diversity and that gave rise to populations outside of Africa through a stepping-stone process.

In the final lab, our discussion focused on gene genealogies and molecular clocks (with the use of DNAsp, ancient human DNA, Neanderthals, and other apes). We discussed the differences between substitution and mutation rates and their relation under the neutral theory (they are the same), how the molecular clock can be calibrated from fossils, and why there might be differences between rates for different loci or lineages, and finally we discussed why there might be problems in estimating fossil age (with ancient DNA available) using the molecular clock (suggested questions for all discussion and assignment can be found in "ESM 2"). Using the tools available in DNAsp (Librado and Rozas 2009), we calculated some population genetic parameters such as genetic diversity (e.g., K and $\pi$), differentiation between populations, and signatures of population expansion and contraction (Tajima's D). This offers a great opportunity to discuss what constitutes a good population sample and why most of our samples do not qualify as unbiased population samples (e.g., human mtDNA clade samples are not a random sample but intended to represent all the diversity observed). Despite bias, using the student population sample (which is closer to a random sample), Tajima's D was significantly negative, indicating a signature of population expansion (as selection would be unlike in a non-coding region of mtDNA). As a summative assessment, students provided written answers to a series of questions related to the analysis of their data; the questions and answers can be found in supplementary information (ESM 2).

## Conclusions

Developing students' tree-thinking skills and a basic understanding of the technical details behind phylogenetic construction is one of the main challenges in evolutionary biology teaching (Baum et al. 2005; Smith and Cheruvelil 2009; O'Hara 1997; Perry et al. 2008). Another major challenge is helping students develop an understanding of gene genealogies (gene trees and species tree differences, different gene genealogies for difference loci) and molecular clocks. The approach presented here has been highly successful in retaining student attention and reinforcing their learning of tree-thinking, phylogenetic reconstruction, population genetics, and molecular clocks. Students reported high satisfaction with the four-week laboratory exercise (100% agreed that the labs were a valuable learning experience and 73% said it was a better educational value than other labs they had in Williams College). Most students did very well on the final assignment, showing both highly developed tree-thinking skills and a deeper understanding of molecular evolution and population genetics. The lab teaching unit described here offers an inquiry-based learning

experience at a relatively low cost per student and gives the opportunity for students to develop skills in many areas of evolutionary biology.

**Acknowledgments** Thanks to Stanislas Monfront for helping to run the first lab trials and to all Williams College students (Bio305) and professors that donated DNA for this lab. Part of this work was carried out using the resources of the Computational Biology Service Unit from Cornell University which is partially funded by Microsoft Corporation.

## References

Baum DA, Offner S. Phylogenies & tree-thinking. Am Biol Teach. 2008;70(4):222–9.

Baum DA, Smith SD, Donovan SSS. Evolution—the tree-thinking challenge. Science. 2005;310(5750):979–80.

Felsenstein J. Inferring phylogenies. Sunderland: Sinauer Associates; 2004. p. 664.

Green RE, et al. A draft sequence of the Neandertal genome. Science. 2010;328(5979):710–22.

Gregory T. Understanding evolutionary trees. Evol: Educ Outreach. 2008;1(2):121–37.

Handelsman J, Miller S, Pfund C. Scientific teaching. New York: W.H. Freeman & Co; 2007. p. 184.

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754–5.

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.

Maddison DR, Maddison WP. MacClade 4: analysis of phylogeny and character evolution. Version 4.08a. . http://macclade.org., 2005.

Meisel R. Teaching tree-thinking to undergraduate biology students. Evol: Educ Outreach. 2010;3(4):621–8.

O'Hara RJ. Population thinking and tree thinking in systematics. Zool Scripta. 1997;26(4):323–9.

Offner S. A universal phylogenetic tree. Am Biol Teach. 2001;63 (3):164–70.

Omland KE, Cook LG, Crisp MD. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. Bioessays. 2008;30(9):854–67.

Perry J, et al. Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. Cbe-Life Sci Educ. 2008;7(2):193–201.

Sandvik H. Tree thinking cannot taken for granted: challenges for teaching phylogenetics. Theory in Biosciences. 2008;127(1):45–51.

Smith J, Cheruvelil K. Using inquiry and tree-thinking to "March Through the Animal Phyla": teaching introductory comparative biology in an evolutionary context. Evol: Educ Outreach. 2009;2 (3):429–44.

Tamura K, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24(8):1596–9.

Tamura K, et al. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.